ASHG 2024 Plenary Abstracts

As of October 17, 2024

Instructions

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

Table of Contents

Session 08: Featured Plenary Abstract Session I

Location: Mile High Ballroom

Session Time: Tuesday, November 5, 2024, 5:00 pm - 6:40 pm

Natural selection shapes the landscape of mosaic chromosomal alterations among participants of diverse ancestries

Authors: M-J. Fave[1], K. Skead[2], V. Bruat[3], J. Kim[2], J. Kang[2], M. Harwood[2], D. Soave[4], H. Noyes[5], P. Awadalla[2]; [1]Concordia Univ., Montréal, QC, Canada, [2]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [3]MaRS Ctr., Toronto, ON, Canada, [4]Wilfrid Laurier Univ., Waterloo, ON, Canada, [5]Univ. of Liverpool, Liverpool, United Kingdom

Abstract:

The age-associated accumulation of somatic mutations in blood, termed Age-Related Clonal Hematopoiesis (ARCH) confers increased risk of morbidity and mortality. Despite significant efforts, the diversity of ancestries in which the genomic, clinical, and environmental associations of ARCH have been reported remains limited, impacting our global ability to generalize findings and undermining our efforts towards genomics equity across human populations. Here we present a combined effort to profile mosaic chromosomal alterations (mCAs) in more than 66,000 samples from diverse ancestries, including 10,500 samples from 7 sub-Saharan and 2 North-African countries, and 30,000 samples from a founder population in Quebec.

We observe the expected increase of mCA prevalence with age, but we concurrently see a decrease in mCAs among individuals starting at around 70 years old, suggesting that clonal expansions in older individuals might reach fixation and are no longer detectable as somatic variants. Interestingly, those trends are observed across ancestries, but with different intensities of decline, suggesting that the evolutionary dynamics within the hematopoietic pool may differ across ancestries or environments. We document replicated and also unique hotspots of somatic mCAs across Africans, Europeans, and the founder-population, and identify ancestry-specific germline variants affecting the risk of developing a clonal expansion. We also document those associations in the founder population in relation to regions of identity-by-descent shared by distantly related participants, which may affect cancer and cardiovascular disease risk differently.

In addition, we reveal transcriptomic effects of clonal expansions in hematopoietic cells on 2000 of our participants using bulk and single-cell RNA-sequencing, and show that negative selection on gene expression plays a key role in regulating the frequency of mCAs in each individual's hematopoietic population. We find that gains, losses, and copy-number-neutral variants impact gene expression distinctly, with stabilizing selection shaping the penetrance of mCAs among blood transcriptomes. Over 30% of canonical eQTLs are impacted by mCAs, revealing a previously overlooked somatic contribution to blood expression profiles, regulatory mechanisms, and disease development.

In summary, we reveal the landscape of mCA across diverse and previously understudied ancestries, we show similar and divergent patterns in the hematopoietic dynamics of aging individuals across populations, and show how gene expression can reveal evolutionary dynamics of clonal expansions within individuals.

Genetic architecture of meiotic recombination across 90,431 *in vitro* fertilized embryos

Authors: R. McCoy[1], A. Biddanda[1], S. Carioscia[1], E. Hoffmann[2]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

Crossovers between homologous chromosomes during meiosis are a key source of human genetic diversity and are also essential for ensuring accurate chromosome segregation. Much of current knowledge about meiotic recombination in the human genome comes from samples of living individuals, thereby solely capturing the minority of crossovers that have survived gametogenesis and fertilization, as well as prenatal and postnatal development. In addition, direct mapping of crossovers based on living human families or gametes has limited analysis to a small number of meioses per individual or a small number of individuals, respectively, hindering understanding of the distribution of variance in crossover phenotypes and its potential genetic determinants. Data from preimplantation genetic testing (PGT) of *in vitro* fertilized (IVF) embryos offer an ideal resource for mapping meiotic crossovers at scale by comparing haplotypes of multiple sibling embryos. Using single nucleotide polymorphism (SNP) array-based PGT data from 46,860 euploid and 43,571 aneuploid blastocyst-stage embryo biopsies across 18,275 sets of parents, we mapped a total of 6,583,309 crossovers across the embryo genomes (3,995,140 maternal, 2,588,169 paternal) to a median resolution of 101 kbp. Leveraging the parental genotype data, we investigated the relationship between parental genetic variation and sex-specific recombination phenotypes across 14,652 paternal and 14,802 maternal genomes. Across six different recombination phenotypes defined by the abundance and locations of recombination events, we identified 34 unique genome-wide significant loci. These include previously reported associations (based on living human trios) at *PRDM9* (linear regression: β; = -0.97, p = 2.8×10^{-275}) for recombination hotspot usage, as well as *RNF212* linear regression: β; = -0.21, p = 1.4×10^{-50}) for total recombination rate. We also identified several novel hits, including common variants in *CCNB1IP1* associated with a reduction in the maternal recombination rate (linear regression: β; = -0.10, p = 1.5×10^{-19}) and a rare variant (0.8% minor allele frequency) in proximity of *DMRT3* associated with an increase in paternal recombination rate (linear regression: β; = 0.45, p = 2.3×10^{-12}). Together, our study showcases the use of PGT data from IVF embryos to clarify variation in the process of meiotic recombination and its complex genetic architecture toward a more complete understanding of this fundamental process shaping human diversity.

Large-scale genomic analysis and targeted functional studies uncover disease-associated uniparental disomy in congenital heart disease

Authors: N. Kong[1], C. Yoon[1], A. Javier[1], W. Dong[2], S. Colijn[1], O. Griffith[1], M. Griffith[1], R. P. Lifton[3], M. Brueckner[2], A. Stratman[1], S. Jin[1], Pediatric Cardiac Genomics Consortium; [1]Washington Univ. Sch. of Med., St. Louis, MO, [2]Yale Univ., New Haven, CT, [3]Rockefeller Univ., New York, NY

Abstract:

Congenital heart disease (CHD) originates from developmental abnormalities in the heart and great vessels, affecting 1-1.8% of live births. Despite extensive research, the etiology of nearly 55% of CHD cases remains unknown. Traditional genetic studies typically identify pathogenic variants by comparing proband's genetics against their parents assuming diploid inheritance. However, this approach fails to detect rare cases of uniparental disomy (UPD), where both chromosomes are inherited from one parent, potentially disrupting genes through loss of heterozygosity and genomic imprinting. This study leveraged data from the Pediatric Cardiac Genomics Consortium, including over 2,750 parent-proband trios with whole-genome sequencing (WGS) and more than 5,350 trios with whole-exome sequencing (WES). Among 4,801 trios analyzed, we identified twelve UPD events, a prevalence 11.2 times greater than that in the general population (two-sided Fisher's $p = 2.5x10^{-9}$). Significant associations were found between CHD and both chromosome 16 UPD and chromosome 4 UPD, as well as maternally inherited UPD overall (two-sided Fisher's p-values: $6.6x10^{-4}$, $3.9x10^{-3}$, and $2.6x10^{-11}$, respectively). To explore the phenotypic impact of UPD, we investigated rare homozygous variants inherited from a single heterozygous parent. This led to the identification of 16 variants of uncertain significance in the UPD regions according to the American College of Medical Genetics and Genomics guidelines. We conducted functional validation using a zebrafish model focusing on the PIEZO1 p.P42L variant found in a patient with coarctation of the aorta. Homozygous loss of piezo1 led to narrowed outflow tracts. The reintroduction of wild-type human PIEZO1 DNA mitigated this phenotype, whereas the p.P42L variant was unable to rescue the phenotype. Additionally, injecting the p.P42L variant into a wild-type zebrafish piezo1 background did not induce phenotypes, indicating the pathogenicity of the p.P42L variant is limited to its homozygous form.
We are actively working to finish additional analyses of WGS and WES data and perform PacBio HiFi sequencing to determine abnormal methylation patterns associated with genomic imprinting of CHD-related genes. Overall, our research bridges significant gaps in understanding the genetic contributors to CHD, with a particular focus on the role of UPD. Elucidating this complex genetic mechanism could improve diagnostics and develop new treatments, ultimately enhancing outcomes for affected individuals.

Long-read RNA-sequencing demarcates *cis*- and *trans*-directed alternative RNA splicing ☆

Authors: G. Quinones-Valdez, K. Amoah, X. Xiao; Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

RNA splicing is influenced by both *cis*-acting genomic elements and *trans*-acting splicing factors, but their specific contributions to splicing events are not well understood. Here, we introduce isoLASER, a computational method that leverages long-read RNA-seq data to classify splicing events as predominantly influenced by either *cis*- or *trans*-directed mechanisms.

IsoLASER computes the adjusted mutual information between local haplotypes and the splicing of exonic segments, identifying allele-specific splicing events with greater sensitivity compared to transcript-level approaches. Its *de novo* variant calling and phasing provide a unique advantage over other methods. Using isoLASER, we identified 2,047 and 4,679 unique exonic regions exhibiting *cis*-directed splicing in multiple human and mouse tissues from ENCODE.

Despite most splicing events being *trans*-directed and tissue-specific, exonic segments with *cis*-directed regulation maintain similar inclusion levels across tissues. Furthermore, genes containing *cis*-directed regulation are enriched in immune-related loci. IsoLASER's HLA-typing function allows for the systematic detection of allele-specific splicing of the HLA gene family, unveiling a new layer of cellular antigen diversification.

Cohort-level analysis, facilitated by isoLASER's joint function, capitalizes on linkage information from multiple donors, revealing splicing-associated variants. Importantly, unlike association studies, this approach does not require a large sample size. When applied to a small cohort consisting of samples from the dorsolateral prefrontal cortex regions of brain tissues from healthy and Alzheimer’s patients, we identified variants linked to splicing in disease-relevant genes such as *MAPT, BIN1,* and *MS4A6A*. Around 7% of the variants identified in this study were categorized as rare or ultra-rare in the population and thus not detectable through QTL mapping methods.

In summary, isoLASER is a novel tool that characterizes both genetically and non-genetically driven splicing events, offering methodological advancements revealing associations independent of cohort size or minor allele frequency. Our findings underscore the potential of isoLASER to enhance our understanding of splicing regulation and its implications in health and disease, particularly within the context of immune response and neurodegenerative disorders

The Biobank Rare Variant consortium powers the discovery of rare genetic associations through global collaboration

Authors: D. Palmer[1,2], M. Brown[3], R. Do[4], C. Gignoux[5], B. Hill[1,2], S. Hodgson[6], G. Kalantzis[7], A. Kousathanas[3], S. Koyama[8,9], R. Loos[4], W. Lu[10,11,12], H. Martin[7], H. Mbarek[13], S. Namba[14,15,16], P. Natarajan[8,17,18,9], Y. Okada[14,15,16], N. Pozdeyev[5,19,20], A. Rendon[3], Z. Rodriguez[21], C. Saad[13], K. Sonehara[14,15,16], D. Van-Heel[22], N. Vartanian[4], A. Verma[21,23,24], H. Vy[4], W. Zhou[10,11,12], C. Lindgren[1,2,25], B. Neale[10,11,12], K. Karczewski[10,11,12], Biobank Rare Variant Analysis consortium; [1]Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom, [2]Nuffield Dept. of Population Hlth., Med. Sci. Div., Univ. of Oxford, Oxford, United Kingdom, [3]Genomics England, London, United Kingdom, [4]Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [6]Queen Mary Univ. of London, London, United Kingdom, [7]Dept. of Human Genetics, Wellcome Sanger Inst., Hinxton, United Kingdom, [8]Ctr. for Genomic Med., Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [9]Cardiovascular Disease Initiative, Broad Inst. of MIT and Harvard, Cambridge, MA, [10]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [11]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [12]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [13]Qatar Genome Program, Qatar Precision Hlth.Inst., Qatar Fndn., Doha, Qatar, [14]Dept. of Genome Informatics, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [15]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Japan, [16]Lab. for Systems Genetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [17]Cardiovascular Res. Ctr., Massachusetts Gen. Hosp., Boston, MA, [18]Dept. of Med., Harvard Med. Sch., Boston, MA, [19]Div. of Endocrinology Metabolism and Diabetes, Dept. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [20]Univ. of Colorado Cancer Ctr., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [21]Dept. of Med., Div. of Translational Med. and Human Genetics, Univ. of Pennsylvania - Perelman Sch. of Med., Philadelphia, PA, [22]Blizard Inst., Queen Mary Univ. of London, London, United Kingdom, [23]Inst. for BioMed. Informatics, Univ. of Pennsylvania - Perelman Sch. of Med., Philadelphia, PA, [24]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [25]Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom

Abstract:

Large-scale biobanks pairing genetic data with deep phenotyping enable the systematic study of genetic influences on diseases and traits. The **B**iobank **Ra**re **Va**riant analysis consortium (BRaVa) has united 16 biobanks and cohorts, comprising > 1.1 million individuals with matched genetic and healthcare record data, to advance the understanding of the genetic basis of human disease.

In our first major BRaVa collaboration, we drew from up to 8 biobanks to conduct rare variant and gene-based meta-analyses across 32 disease endpoints and 10 biomarkers and continuous traits. These analyses incorporated individuals with diverse ancestries, including over 17,000 East Asians, 50,000 South/Central Asians, 70,000 Africans, 44,000 Admixed Americans, and 645,000 Europeans. We harmonized phenotype definitions and analyzed both common diseases such as atrial fibrillation (57,000 cases, 748,000 controls) and less prevalent conditions such as infertility (female infertility; 5,000 cases, 399,000 controls), as well as anthropometric traits including BMI (N = 696,000). In addition to discovering novel associations, BRaVa is also addressing common challenges in biobank-based genetic studies, including best practices for quality control of genetic and phenotypic data as well as variant annotation and prioritization.

All-biobank gene-based meta-analyses of rare (MAF < 0.001) putatively damaging classes of variation across 42 traits identified 279 (89 binary, 181 continuous) unique gene-trait associations (experiment-wise $P < 2.5 \times 10^{-7}$, based on effective number of tests via simulation), of which 131 are potentially novel, i.e. have not been observed in large biobank sequencing studies. Among these, 8 genes showed differential burden effect sizes across biobanks. Exemplar associations include that between damaging variation in *ANKRD12* and increased risk of chronic obstructive pulmonary disorder, as well as burden of ultra-rare (MAF < 0.0001) predicted loss-of-function variation in *KDM6B* and female infertility.

Despite differences in phenotype ascertainment and sample recruitment across biobanks, the increased sample size garnered by BRaVa boosted power to detect rare genetic associations. E.g., among disease endpoints, highly significant ($P < 1 \times 10^{-10}$) rare predicted loss-of-function gene-based associations, the median ratio between χ^2 statistics from our meta-analysis and the European subset of the UK Biobank was 1.8, compared to the median ratio between effective sample sizes of 2.0.
These analyses represent the world's largest rare-variant association study (RVAS), and provide a roadmap to RVAS meta-analysis with millions of individuals.

Session 60: Featured Plenary Abstract Session II

Location: Mile High Ballroom

Session Time: Thursday, November 7, 2024, 5:00 pm - 6:40 pm

Dissecting Genetic Heritability, Environmental Risk, and Causal Effects of Air Pollution Using a Health Insurance Database of >50 Million Individuals ☆

Authors: H. Markus[1], D. McGuire[1], L. Yang[1], J. Xu[1], A. Montgomery[1], A. Berg[1], Q. Li[2], L. Carrel[1], D. Liu[1], B. Jiang[1]; [1]Penn State Coll. of Med., Hershey, PA, [2]Penn State Univ., University Park, PA

Abstract:

Most complex diseases are jointly influenced by both genetics and environment. The advent of large national-level electronic health record (EHR) datasets has offered new opportunities for disentangling the role of genes and environment through the deep phenotype information and approximate pedigree structures that EHR datasets provide. In this study, we made use of the approximate geographical locations of patients as a proxy for spatially correlated community-level environmental risk factors and developed a spatial mixed linear effect (SMILE) model that incorporates both genetics and environmental contribution. The SMILE model improves the estimates of heritability, as community-level environment (such as air pollution) is often shared across families living in similar locations but are typically insufficiently considered in traditional family-based variance components models, leading to inflated estimates of genetic heritability.

To apply the SMILE model, we extracted EHR and approximate geographical locations from 257,620 nuclear families and compiled 1,083 disease outcome measurements from the MarketScan dataset. We further augmented the EHR with publicly available environmental data, including levels of particulate matter 2.5 ($PM_{2.5}$), nitrogen dioxide (NO_2), climate, and sociodemographic data. We refined the estimates of genetic heritability and quantified community-level environmental contributions. We also used wind speed and direction as instrumental variables in the SMILE model to assess the causal effects of air pollution. In total, we found $PM_{2.5}$ or NO_2 have statistically significant putative causal effects on 135 diseases, including respiratory, musculoskeletal, digestive, metabolic, and sleep disorders. Many of these associations have previously implicated plausible biological mechanisms, although some were reported only in smaller studies from heavily polluted areas. While individual air pollutant levels often stem from common sources such as traffic pollution, we found $PM_{2.5}$ and NO_2 to have unique disease etiologies and affect biologically distinct disease categories. These analyses showcase several novel strategies for jointly modeling genetic and environmental effects on disease risk using large EHR datasets and will benefit upcoming biobank studies in the era of precision medicine.

Methylome and epigenome dynamics during early colorectal oncogenesis

Authors: H. Lee[1], G. Krieger[1], T. Clark[2], A. Khan[3], Y. Zhu[3], E. Esplin[4], C. Hanson[3], A. Horning[3], M. babu[3], K. Paul[3], R. Chiu[3], J. Bregman[5], E. Griffin-Baldwin[6], I. Rao[1], Y. Agarwal[1], B. Bahmani[1], S. Nevins[3], T. Karathanos[3], A. Weimer[3], E. Meiri[2], S. Gilad[2], S. Benjamin[2], D. Lebanony[2], N. Iremadze[2], F. Oberstrass[2], A. Jaimovich[7], A. Bellacosa[1], J. Ford[3], W. Greenleaf[3], D. Lipson[2], Z. Shipony[8], M. Snyder[3]; [1]Fox Chase Cancer Ctr., Philadelphia, PA, [2]Ultima Genomics, Fremont, CA, [3]Stanford Univ., Stanford, CA, [4]Invitae, San Francisco, CA, [5]Temple Univ., Philadelphia, PA, [6]Southwestern Univeristy, Georgetown, TX, [7]Ultima Genomics, Hertzliya, CA, [8]Ultima Genomics, Rehovot, Israel

Abstract:

Alterations in DNA methylation have been associated with cancer onset and progression. However, it is not clear how early methylation changes occur and how they interact with other epigenomic shifts. To address epigenomic dynamics during the early stages of transformation from normal healthy colon tissue to early cell growth and malignant colorectal cancer (CRC), we collected 51 samples of normal colonic mucosa, benign and dysplastic polyps and adenocarcinoma from 15 familial adenomatous polyposis (FAP) patients and non-FAP colorectal cancer patients. We generated 30-100x (60x average) whole-genome enzymatic methylation sequencing (WGEM-seq) using the novel Ultima Genomics ultra high-throughput sequencing platform. We observed that hypomethylation emerges throughout the genome, including at centromeric and subtelomeric regions, early in the transformation to malignancy. Hypermethylation is prevalent (>80%) in the cell cycle regulator gene CCNA1 and in the chromatin regulator gene KDM2B. Interestingly, epigenetic age undergoes significant acceleration during the transition from normal colonic mucosa to benign polyps, averaging 15-20 years, indicating the potential of epigenetic age acceleration as an early biomarker for detecting colorectal cancer. Intra-patient polyp samples reveal a coexistence of stemness and aging, highlighting the antagonistic biological processes occurring within polyps. We integrated methylation alterations with whole-genome sequencing (WGS), chromatin accessibility (ATAC-seq), fine-resolution long-range chromatin conformation capture (Tri-C), and gene expression (RNA-seq) data to investigate how epigenomic and genomic alterations collectively affected gene expression. Our prediction model, designed to capture interactions among diverse epigenomic elements, revealed that promoter methylation has the most significant opposing impact on gene expression, followed by single nucleotide variations (SNVs), which showed the second highest anti-correlation. Surprisingly, methylation in promoters with polycomb repressive complex2 (PRC2) binding sites exhibited a positive correlation with gene expression. This study reveals that aberrant methylation occurs earlier and is more prevalent than other genomic and epigenomic alterations, suggesting its driving role in early oncogenesis and supporting its effectiveness as a biomarker for early diagnosis and potential for targeted therapeutics.

Integrative -omics analysis of endometriosis uncovers key pathways from a multi-ancestry study of over 800,000 women ☆

Authors: L. Guare[1], J. Das[1], E. Moreno[2], A. J. Mulford[3], B. Brumpton[2], S. Chapman[4], T-T. Chen[5], Y-F. Lin[6], A. Sanders[3,7], Y. Shi[8], Y. Shirai[9], J. Shortt[10], D. Velez Edwards[11], W. Zhou[12], N. Elhadad[13], G. Jarvik[14], L. Kottyan[15], Y. Luo[16], W-Q. Wei[17], C. Weng[18], Penn Medicine Biobank, Global Biobank Meta-Analysis Initiative, Regeneron Genetics Center, S. Verma[19]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [3]Endeavor Hlth., Evanston, IL, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Natl. Hlth.Res. Inst.s, Taiwan, Miaoli, Taiwan, [6]Natl. Hlth.Res. Inst.s, Taiwan, Zhunan Town, Miaoli County, China, [7]Univ. of Chicago, Chicago, IL, [8]Broad Inst., Cambridge, MA, [9]Osaka Univ., Suita, Japan, [10]Colorado Ctr. for Personalized Med., Aurora, CO, [11]Vanderbilt Univ Med Ctr., Nashville, TN, [12]Massachusetts Gen. Hosp., Chestnut Hill, MA, [13]Columbia Univ., New York City, NY, [14]Univ Washington Med Ctr., Seattle, WA, [15]Cincinnati Children s Hosp. Med. Ctr., Cincinnati, OH, [16]Northwestern Univ. Feinberg Sch. of Med., Evanston, IL, [17]Vanderbilt Univ, Nashville, TN, [18]Columbia, New York City, NY, [19]Univ. of Pennsylvania, Cherry Hill, NJ

Abstract:

Endometriosis (endo) is a heritable disorder (twin h2 ~ 47%) characterized by infertility and severe pain affecting up to 10% of women of reproductive age. Multi-systemic, heterogeneous symptoms delay average diagnosis to 5 years or more. Despite the high prevalence and severe symptoms, the precise disease mechanisms of endo remain under-studied. Previous GWASs have been performed in predominantly European-ancestry (EUR) cohorts. We performed a genome-wide meta-analysis for endometriosis consisting of a large portion (35%) of non-EUR samples across 12 biobanks as part of Global BioBank Meta-Analyses Initiative. We defined wide and narrow phenotypes using EHRs for our study: wide endo (W), wide endo excluding adenomyosis (Wex), and procedure-confirmed or surgically confirmed narrow endo (PCN and SCN). GWASs for the phenotypes were run by the collaborating biobanks (AoU, BBJ, BioVU, eMERGE, FinnGen, GHI, HUNT, Latvian National Biobank, MGBB, PMBB, TWBB, and UKBB) and subsequently meta-analyzed. Post-GWAS, we performed heritability estimation, fine-mapping, TWAS, PWAS, pathway enrichment testing, and integrative multi-omics analyses. Our largest analysis consisted of 39,764 cases (10,871 non-EUR) and 784,147 controls (246,040 non-EUR), totaling 823,911 women. We identified 43 significant loci, five of which are previously unreported: *EEFSEC* (rs2955117, P=1.88E-10), *DTD1* (rs6081317, P=9.79E-10), *LOC100131685* (rs6778588, P=9.98E-10), *ZHX3* (rs17181845, P=2.99E-09), and *MIRLET7BHG* (rs8136639, P=2.98E-08). These novel GWAS loci are previously linked to endo risk factors such as age at menarche and smoking. Our Wex analysis (N = 380,966) replicated the *SKAP1* locus (rs112577355, P=3.22E-9). Significant ancestry-stratified heritability estimates were 11.5% for EUR, 10.6% for EAS and 27.3% for AFR. 42 total causal SNPs were found in the meta-analyses: 32 multi-ancestry, 1 AFR, and 9 EUR. TWAS identified 14 associated genes, while PWAS suggests significant association of R-spondin 3 (RSPO3) that plays a crucial role in modulating the WNT signaling pathway. Our diverse, comprehensive GWAS, coupled with integrative omics analysis, identifies critical pathways in endometriosis pathogenesis: DDX58/IFIH1 pathway had been previously reported in immunopathogenesis of endo, WNT signaling followed by MET-activated PTK2 and RhoC GTPase implicates importance of critical balance between proliferation, differentiation, and migration of endometrial cells. These interconnected pathways and risk factors underscore a complex, multi-faceted etiology of endo, presenting multiple targets for precise and effective therapeutic interventions.

Joint profiling of Cas9 edit sites and transcriptomes in single-cells with Superb-seq

Authors: G. McVicker[1], M. Lorenzini[1,2], K. Sajeev[2], B. Balderson[1], A. Ho[1]; [1]Salk Inst. for Biological Studies, La Jolla, CA, [2]Univ. of California San Diego, La Jolla, CA

Abstract:

A current barrier in CRISPR-Cas9 pooled screens is the inability to directly measure on- and off-target edits and the effects of both kinds of edits on gene expression. Here, we present a single-cell edit capture sequencing (Superb-seq) technology to jointly profile Cas9 edits and transcriptomes by single-cell RNA-seq. The Superb-seq workflow consists of three steps. First, we label Cas9 cleavage events with phage T7 promoter sequences through live-cell electroporation of double-stranded oligodeoxynucleotides (dsODNs). Second, we fix cells and perform T7 in situ transcription to generate T7 RNAs. Third, we perform combinatorial barcoding and sequencing of in situ T7 transcripts and endogenous transcriptomes with off-the-shelf SPLiT-seq. Superb-seq achieves 40-80% edit labeling at six tested loci in three human cell types, including primary T cells. We performed Superb-seq on 9,500 K562 cells, targeting four chromatin remodeler genes using seven guide RNAs. Using the data generated by this experiment, we demonstrate (1) direct single-cell capture of on-target edit events, (2) high-throughput single-cell detection of Cas9 off-target events, (3) single-cell edit allele counting, and (4) associations between edit events and gene expression. The simple Superb-seq protocol uses commercial kits, standard laboratory equipment, and requires no virus, enabling any laboratory to perform these experiments. We envision that Superb-seq will be immediately useful for (1) targeted pooled and in vivo ribonucleoprotein CRISPR screens in hard-to-transduce cell types, and (2) detection of off-target perturbations. Future scaling of Superb-seq with pooled guides will enable powerful Perturb-seq screens and validation of widely used guide RNA libraries.

Transcriptomic-first approach for rare disease diagnostics uncovers global dysregulation of splicing caused by minor spliceosome core gene mutations and identifies candidate novel gene-disease relationship ☆

Authors: M. Maurer[1,2,3], R. Mendez[3,4], V. S. Ganesh[5,6,7], G. Lemire[5,7], R. Ungar[1,2,3], D. Bonner[2,3,8], D. R. Matalon[8], J. A. Bernstein[3,8], Undiagnosed Diseases Network, A. O'Donnell-Luria[5,7], M. T. Wheeler[1,3,4], S. B. Montgomery[1,2,3]; [1]Dept. of Genetics, Stanford Univ., Stanford, CA, [2]Dept. of Pathology, Stanford Univ., Stanford, CA, [3]Stanford Ctr. for Undiagnosed Diseases, Stanford, CA, [4]Dept of Cardiology, Stanford Univ., Stanford, CA, [5]Ctr. for Mendelian Genomics, Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Dept. of Neurology, Brigham and Women's Hosp, Boston, MA, [7]Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, [8]Div. of Med. Genetics, Dept. of Pediatrics, Stanford Univ., Stanford, CA

Abstract:

Integration of RNA-seq data has proven useful for increasing the diagnostic yield in rare disease patients. Current approaches focus on identifying variants resulting in single transcriptomic outlier events (i.e., over or under expression); however, this myopic approach ignores broader transcriptomic patterns that could infer a specific causal monogenic mechanism. Here, we present a transcriptomics-first method to diagnose rare disease patients by examining transcriptome-wide patterns of minor intron retention events. Minor introns represent ~0.5% of all human introns and are spliced by separate machinery: the minor spliceosome, which consists of five noncoding RNAs and over 170 proteins.

Using pre-existing splicing outlier detection methods, FRASER and LeafcutterMD, we called splicing outliers in whole blood samples from 210 rare diseases probands and 185 familial controls from the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) and Undiagnosed Diseases Network consortiums. While evaluating global patterns of splicing outliers, we utilized the Minor Intron Database to compute the number of minor intron containing genes (MIGs) with intron retention events per sample.

The mean number of MIGs affected by intron retention events is 4.3 per sample (± 26.2). We observed that three unrelated probands had a considerable excess of MIGs with intron retention events (z-scores of 12.0, 10.7, and 10.1, corresponding to 323, 289, and 271 out of 699 possible MIGs impacted respectively). Two of these individuals were found to harbor biallelic pathogenic variants in the noncoding minor-spliceosome gene *RNU4ATAC*. Both patients shared features of known RNU4ATAC-opathies, such as intrauterine growth retardation (IUGR), short stature, neurodevelopmental delay, retinal dystrophy, and immunodeficiency. Genomic analysis of the third proband - a child with IUGR, microcephaly, short stature, epilepsy, intellectual disability, and ataxia - revealed rare compound heterozygous variants of uncertain significance in the noncoding minor-spliceosome gene *RNU6ATAC*. Our genomic and phenotypic findings strongly suggest *RNU6ATAC* as a novel gene-disease candidate.

By examining patterns of transcriptome-wide splicing perturbations, we confirmed two diagnoses of RNU4ATAC-opathies and uncovered a candidate gene-disease relationship. Outside of MIGs, we uncovered 23 additional probands with excess numbers of splicing outliers. Dissecting transcriptome-wide patterns of aberrant splicing may represent a novel diagnostic approach for gene discovery and variant-to-functional interpretation of splicesomopathies in rare disease patients.

Session 86: Featured Plenary Abstract Session III

Location: Mile High Ballroom

Session Time: Friday, November 8, 2024, 5:00 pm - 6:40 pm

Inference of genome-wide genealogical relationships between ancient and modern individuals ☆

Authors: Z. Tsangalidou[1], J. Sant[1], A. Gunnarsson[2,1], R. Fournier[1], P. Palamara[1,3]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]deCODE genetics, Reykjavík, Iceland, [3]Ctr. for Human Genetics, Oxford, United Kingdom

Abstract:

Recent advances in sequencing technologies have greatly increased the number of available ancient DNA (aDNA) human genomes. Combining aDNA with large modern biobanks can provide new insights into the relationships between ancient and modern groups and reveal details of human evolutionary history. We developed a method, ThreaDNA, to study the genome-wide genealogical relationships between ancient and modern individuals by threading the unphased diploid genome of ancient individuals into an ancestral recombination graph (ARG) inferred from modern data. This approach relies on an efficient haplotype matching algorithm based on the Li-Stephens model and the positional Burrows-Wheeler transform, coupled with approximate likelihood-based inference of coalescence time.

Using simulations, we verified that ThreaDNA is accurate when applied to unphased aDNA samples genotyped at a coverage of 0.5x or higher and scales linearly with the number of ancient and modern individuals. We then used it to thread diverse imputed ancient individuals from the Allen Ancient DNA Resource dataset to a genome-wide ARG comprising 487,409 modern UK Biobank samples, which we inferred by applying the Threads algorithm to 9,992,478 high-quality imputed variants.

Analyzing the genealogical connections within this inferred ARG, which jointly models both modern and ancient individuals, we recovered demographic shifts within the past ~2000 years that reflect recent historical events in the UK. We observed an increase in shared recent ancestry between Viking samples and UK Biobank samples from the north and south of England, compared to those from Scotland and Wales. We also found that Norwegian Viking samples were more closely linked to the Hebrides and Orkney regions compared to Danish Viking samples, aligning with historical migration patterns. In addition, we found ancestry derived from ancient Irish individuals to correlate with Gaelic-speaking areas of Scotland and Northern Ireland. Roman and Iron Age samples were most closely related to areas in Wales, Scotland, and north-west England, complementary to those involving Saxon migrations into the UK. Finally, we created a genome-wide annotation capturing recent ancestry from various Viking, Saxon, British, and Irish ancient groups, finding this annotation to be enriched for known selection targets, including the LCT and HLA regions. These findings underscore the utility of constructing large-scale genealogies that include both ancient and modern individuals to gain insights into historical migrations, population dynamics, and recent natural selection within the UK.

Revealing the exact breakpoints and sequence rearrangements of recurrent, large neuropsychiatric copy number variations (CNVs) at single base-pair resolution using CRISPR-targeted ultra-long read sequencing (CTLR-Seq)

Authors: A. Urban[1], B. Zhou[2], C. Purmann[1], H. Guo[1], G. Shin[1], Y. Huang[1], Q. Meng[1], R. Pattni[1], S. Greer[1], T. Roychowdhury[3], R. N. Wood[1], M. Ho[1], H. B. zu Dohna-Schlohbitten[4], A. Abyzov[5], J. Hallmayer[1], W. Wong[1], H. Ji[1]; [1]Stanford Univ., Palo Alto, CA, [2]Stanford Univ. Sch. of Med., Palo Alto, CA, [3]Yale Univ., New Haven, CT, [4]American Univ. of Beirut, Beirut, Lebanon, [5]Mayo Clinic, Rochester, MN

Abstract:

Here we report for the first time the exact breakpoints and structures of the breakpoint regions for the recurring neuropsychiatric large copy number variants (CNVs), using a novel method combining in vitro CRISPR and long-read sequencing.

The large CNVs in 16p11, 22q11, 1q21, 15q13, and 3q29 are recurrent heterozygous chromosomal aberrations that are the highest known, relatively frequent, genetic risk factors for psychiatric disorders such as schizophrenia and autism. They are flanked by Segmental duplications (SegDups), extended, complex, repeating cassettes of genomic sequence.

These SegDups have so far been impenetrable to full analysis including by standard whole-genome sequencing. We developed CTLR-Seq, combining in vitro CRISPR-cutting and ultra-long nanopore sequencing, to isolate, sequence, assemble, and resolve very large CNV rearrangements regions (often around 700 kbp) on the deletion haplotype.

With CTLR-Seq we now resolved at nucleotide level the major neuropsychiatric CNVs: 16p11.2 deletion (n=7),16p11.2 duplication (n=12), 15q13.3 deletion (n=4), 15q13.3 duplication (n=3), 1q21 deletion (n=4), 22q11.2 (typical 3Mb) deletion (n=5), and 3q29 deletion (n=1), in each case resolving CNV flanking regions several hundred kbp in length.

All 22q11 deletions analyzed contain a retrotransposon at the breakpoint, suggesting formation via transposon-mediated recombination. Retrotransposons also reside at the breakpoints of other major large recurrent CNVs. We demonstrate that CTLR-Seq results now enable functional genomics analysis within the SegDup regions. We performed chromosomal folding analysis with HiChIP within the SegDups in neurons and astrocytes from 22q11.2 deletion patient-derived neural organoids. Chromosome interactions anchored within the SegDups are both cell type-specific and patient-specific. Further, we demonstrate that CTLR-Seq enables cell-type specific analysis of DNA methylation patterns within the SegDup sequences of the deletion haplotype of 22q11.2.

CTLR-Seq allows for complete resolution of large sequence elements such as SegDups that are inaccessible with other methods. We resolve the breakpoint regions of the major neuropsychiatric CNVs and demonstrate that current standard analysis of their extent is uncertain by up to several hundred kbp. Such information may enable better understanding of the mechanisms of formation of such CNVs, comprehensive analysis of the functional genomics effects, and eventually clinical genetics determination of the exact nature of such CNVs in individual patients, also in the context of understanding individual risk and variance of symptoms.

Variation in chromatin accessibility in immune cells: population differences and genetic contributors revealed by single-cell approaches

Authors: Y. Aquino[1], A. Bisiaux[1], Z. Li[1], M. Hasan[1], E. Patin[1], M. Rotival[1], L. Quintana-Murci[1,2]; [1]Inst. Pasteur, Paris, France, [2]Coll. de France, Paris, France

Abstract:

Transcriptional responses to infection vary widely across individuals and populations. While growing evidence links differences in immune responses with epigenetic variation, the specific mechanisms through which chromatin remodeling contributes to the diversity of transcriptional responses to immune stimulation remain elusive. Here, we used single-nuclei assays of transposase-accessible-chromatin sequencing (snATAC-seq) on peripheral blood mononuclear cells from 160 healthy donors of Central African or West European ancestry to delineate the chromatin accessibility landscape across 21 immune cell types. We uncovered 2,197 open chromatin regions (OCRs) displaying differential accessibility ($|\log_2 FC| \geq 0.5$, FDR ≤ 0.01) between the two populations (popDA), with κ-light-chain memory B cells being the most differentiated (1,013 popDA OCRs). Interestingly, around 63% of these popDA OCRs are more accessible in African-ancestry memory B cells and enriched in binding motifs for T-box transcription factors involved in the acquisition of immune memory, like T-bet (*TBX21*). This suggests a preferential switch of memory B cells to T-bet+ phenotypes in individuals of African ancestry. We next mapped the genetic determinants of chromatin accessibility, using RASQUAL, and identified 20,174 OCRs associated (FDR < 0.05) with nearby genetic variants (cis-caQTL). Notably, 80% of caQTL were detected in a single immune lineage, highlighting the strong cell context-dependent nature of the genetic determinants of chromatin accessibility. Focusing on population variation, popDA OCRs were enriched in cis-caQTL in all immune lineages (OR > 3.2, $p < 2.2\times10^{-16}$). However, only around 3% of T-bet motif-harboring OCRs displayed a significant caQTL in B cells, pointing to environmental or trans-genetic effects as sources of the African T-bet+ phenotype. Finally, leveraging scRNA-seq data from the same individuals, at basal state and after SARS-CoV-2 stimulation, we estimated that only up to 5% of caQTL are also associated with gene expression (eQTLs, $p < 10^{-3}$), suggesting that caQTL and eQTL mapping studies target fundamentally different sets of regulatory variants. In line with these observations, we found that while caQTLs are enriched among variants associated to immune traits genome-wide (OR = 4.07, $p < 2.2\times10^{-16}$), over half of these signals (88/169) were missed by previous eQTL mapping. Overall, our results highlight the relevance of mapping caQTL to delineate a potential missing layer of regulation that could enhance the prediction of complex disease risk from genome-wide associations.

Quantifying the impact of genetic variation on healthcare cost across 1.5 million individuals from 13 studies and 8 countries: the GenCost consortium ☆

Authors: S. May-Wilson[1], T. Nakanishi[1], J. Lee[2], I. Louloudis[3], P. A. Lind[4], A. Richmond[5], C. Saad[6], C. M. van der Laan[7], S. Kanoni[8], P. Fahr[9], A. Kukkonen[10], K. Zguro[1], Z. Yang[1], P. M. Torkki[1], E. Abner[10], P. Natarajan[11], FinnGen Consortium, d. van heel[12], H. Mbarek[13], D. Boomsma[14], R. Marioni[5], N. G. Martin[15], N. Davies[16], S. E. Medland[17], S. Brunak[3], P. Dixon[9], A. Ganna[18]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Broad Inst. of MIT and Harvard, Finnish Inst. for Molecular Med., Cambridge, MA, [3]Univ. of Copenhagen, Copenhagen, Denmark, [4]QIMR Berghofer Med. Res. Inst., Herston, Australia, [5]Univ. of Edinburgh, Edinburgh, United Kingdom, [6]Qatar Precision Hlth.Inst., Doha, Qatar, [7]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [8]Queen Mary Univ. of London, London, United Kingdom, [9]Univ. of Oxford, Oxford, United Kingdom, [10]Univ. of Tartu, Tartu, Estonia, [11]Massachusetts Gen. Hosp., Boston, MA, [12]Barts and The London, London, United Kingdom, [13]Qatar Genome, Doha, Qatar, [14]VU Univ. Amsterdam, Amsterdam, Netherlands, [15]Queensland Inst Med Res, Brisbane, Australia, [16]Univ. Coll. London, London, United Kingdom, [17]QIMR Berghofer, Brisbane, Australia, [18]Inst. for Molecular Med. Finland, Helsinki, Finland

Abstract:

An individual's healthcare costs can provide an objective measure of overall morbidity. Quantifying the impact of genetic variation on healthcare costs can inform public health policies and help identify genetic determinants of mortality.

We estimated inpatient, outpatient, primary care and prescription-related costs, from the perspective of healthcare systems, for a total of 1.45 million individuals across 8 different healthcare systems. The median per-patient annual healthcare cost (expressed in 2019 prices) had substantial range across countries, from €483/year in the Estonian Biobank to €458,455/year in the Mass General Brigham Biobank. A GWAS meta-analysis identified 27 and 106 significant, independent loci for inpatient and medication costs, respectively. Effects were overall modest, the most significant inpatient signal (rs6938239 near *ILRUN*) was associated with ~1% increased cost per year, per allele. Following stratified analyses, 33% of inpatient loci had sex-specific effects while 22% had an age-specific effect. E.g. homozygous carriers of variant rs62236881 in the 3' UTR of *ZNRF3* had 9% higher costs in females compared with 4% in males, potentially due to an increased risk of breast cancer.

Despite differences in tariff structure and cost reimbursement across healthcare systems, we observed strong genetic correlations between inpatient studies (median rg=0.88) and 95% of significant loci did not show heterogeneity across studies. PheWAS indicated that significant loci may impact healthcare costs via either common risk factors (e.g. BMI) or chronic diseases risk (e.g. asthma, depression).

Leveraging whole-exome sequencing data from the UK Biobank (UKB), we estimated the cost for putative loss of function (pLOF) burdens for 83 clinically relevant genes. The largest effects were observed for *BRCA1/2*, *MSH2* and *APC*. Annual inpatient costs were ~1.8 times higher in individuals carrying one pLOF in *BRCA2* than in non-carriers. We estimated that the total annual inpatient costs attributed to *BRCA2* pLOF mutations amongst those aged 40-69 in the UK would be about €26 million.

Polygenic scores (PGS) for various risk factors and diseases were linked to higher healthcare costs. In UKB, the top 10% with the highest BMI PGS had a median annual cost of €414, compared to €329 for those in the middle 40-60%.

We derived a PGS for in-patient cost a tested in ~11,000 Genomics England participants showing the top 10% incurred ~€250/year more than those in the middle 40-60%.

Genetic variation across individuals influences healthcare expenditure. Our findings are applicable across various healthcare systems and help better understanding global morbidity.

Expanded Newborn Screening Using Genome Sequencing for Early Actionable Conditions: results of the first 10,000 participants enrolled in the GUARDIAN study

Authors: W. Chung[1], M. Jeanne[1], A. Ziegler[2], C. Koval-Burt[2], D. M. Kay[3], S. F. Suchy[4], A. Begtrup[4], K. Langley[4], L. Amendola[5], B. M. Boyd[2], T. Brandt[6], L. L. Cohen[7], A. Coffey[8], J. M. Devaney[4], B. Friedman[6], R. Hernan[1], S. Hofherr[4], K. S. Hruska[4], Z. Hu[9], A. Johnson[4], R. L. Leibel[2], K. McWalter[4], K. G. Monaghan[4], K. Retterer[4], R. J. Sicko[3], S. Strom[10], R. I. Torene[4], R. J. Taft[5], P. Kruszka[4], M. Caggana[11]; [1]Dept. of Pediatrics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [2]Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, [3]Newborn Screening Program, Wadsworth Ctr., New York State Dept. of Hlth., Albany, NY, [4]GeneDx, Gaithersburg, MD, [5]Illumina, Inc, San Diego, CA, [6]Gene Dx, Gaithersburg, MD, [7]Dept. of Pediatrics, Weill Cornell Med. Coll., New York, NY, [8]Illumina Inc, San Diego, CA, [9]Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, [10]Illumina Inc., San Diego, CA, [11]New York State Dept Hlth, Albany, NY

Abstract:

Introduction: The use of genetic sequencing in current Newborn Screening (NBS) programs is restricted to second-tier testing, but improved interpretation of genomic data and declining cost of DNA sequencing have raised the question of its implementation in first-tier testing. Genomic NBS could allow states to expand screening for conditions lacking biomarkers and increase sensitivity and specificity of standard NBS. However, several challenges need to be addressed such as equity, scalability, and variant interpretation across diverse ancestral groups.
Methods: Genomic Uniform Screening Against Rare Disease in All Newborns (GUARDIAN) is a prospective pilot study investigating the use of targeted interrogation of the genome as part of routine NBS. Dried blood spots collected as part of routine NBS are used for genome sequencing. A total of 156 early-onset and treatable conditions (group 1) were screened in all participants from September 2022 to February 2024. Parents were also offered the option for additional screening for 99 neurodevelopmental disorders (group 2). Since March 2024, the number of conditions screened was increased to 321 in group 1 and 144 in group 2.
Results: Over 18 months, 13,466 families were approached, and 10,000 (74.0%) consented to participate. Enrolled participants represented a diverse group by race and ethnicity, mirroring the diversity found in New York City. A majority of parents consented for screening of both group 1 and 2 (92.6%). The positive screen rate was 3.3%, with a positive predictive value of 79% across all conditions. DNA sequencing improved standard NBS by identifying false negatives for 2 cases of severe combined immunodeficiency. Treatable conditions that are not currently included in NBS included 11 cases of long QT syndrome, most associated with a prolonged QT interval on ECG in the baby. There were 16 cases with pathogenic/likely pathogenic variants in *SCN1A,* 2 de novo, 12 inherited, and 2 of unknown inheritance; 1 associated with neurodevelopmental delay, 1 in an infant with febrile seizures, and 2 associated with seizures in a heterozygous parent. For recessive conditions, phasing with parental testing identified 15 variants in cis that were false positive. Predictions of phase using gnomAD co-occurrence data accurately predicted phase for 27/28 variant pairs assessed.
Conclusions: Our findings demonstrate the feasibility of screening for a targeted set of genes in a diverse newborn population. Additional studies are required to assess long-term outcomes, barriers to implementation, cost-effectiveness, and to provide guidelines for its potential use for routine NBS.